Computer-Aided Design of Antimicrobial Peptides

Fjell, Christopher D.; Hancock, Robert E.W.; Jenssen, Håvard

Published in:
Current Pharmaceutical Analysis

DOI:
10.2174/157341210791202645

Publication date:
2010

Document Version
Early version, also known as pre-print

Citation for published version (APA):
Fjell, C. D., Hancock, R. E. W., & Jenssen, H. (2010). Computer-Aided Design of Antimicrobial Peptides. *Current Pharmaceutical Analysis*, *6*(2), 66-75. https://doi.org/10.2174/157341210791202645

Computer-Aided Design of Antimicrobial Peptides

Christopher D. Fjell[1], Robert E.W. Hancock[1] and Håvard Jenssen[1,2,*]

[1]*Centre for Microbial Diseases and Immunity Research, University of British Columbia, 2259 Lower Mall Research Station, Vancouver, British Columbia, V6T 1Z3, Canada*

[2]*Roskilde University, Dept. of Science, Systems & Models, Universitetsvej 1, Building 18.1, DK-4000 Roskilde, Denmark*

Abstract: An increasing number of reported cases of drug resistant *Staphylococcus aureus* and *Pseudomonas aeruginosa*, demonstrate the urgent need for new therapeutics that are effective against such and other multi-drug resistant bacteria. Antimicrobial peptides have for two decades now been looked upon as interesting leads for development of new therapeutics combating these drug resistant microbes.

High-throughput screening of peptide libraries have generated large amounts of information on peptide activities. However, scientists still struggle with explaining the specific peptide motifs resulting in antimicrobial activity. Consequently, the majority of peptides put into clinical trials have failed at some point, underlining the importance of a thorough peptide optimization.

An important tool in peptide design and optimization is quantitative structure-activity relationship (QSAR) analysis, correlating chemical parameters with biological activities of the peptide, using statistical methods. In this review we will discuss two different *in silico* strategies of computer-aided antibacterial peptide design, a linear correlation model build as an extension of traditional principal component analysis (PCA) and a non-linear artificial neural network model. Studies on structurally diverse peptides, have concluded that the PCA derived model are able to guide the antibacterial peptide design in a meaningful way, however requiring rather a high homology between the peptides in the test-set and the *in silico* library, to ensure a successful prediction. In contrast, the neural network model, though significantly less explored in relation to antimicrobial peptide design, has proven extremely promising, demonstrating impressive prediction success and ranking of random peptide libraries correlating well with measured activities.

Keywords: *P. aeruginosa*, Antimicrobial Peptides, Quantitative Structure-Activity Relationships, Prediction of activity, Partial Least Square Projections to Latent Structures, Artificial Neural Network.

INTRODUCTION

The spread of antibiotic-resistance amongst bacterial pathogens, has resulted in an dramatic increase in methicillin-resistant *Staphylococcus aureus* (MRSA) [1,2] and *Streptococcus pneumonae* resistance [3]. Despite these alarming trends and obvious need for new interventions, pharmaceutical companies have withdrawn from the field of anti-infectives [4], introducing only two novel antibiotics to the market in the last 20 years [5].

Cationic antimicrobial peptides, also called cationic host defence peptides, are signature antimicrobials of nearly all species of life, and their broad spectrum activity and rapid action [6,7] render them as one of the most fascinating candidates for new antibiotics [8,9]. More than 1000 natural occurring peptides have been described so far, and the majority of these are described in databases for eukaryotic host defence peptides: Alessandro Tossi's site at the University of Trieste (http://www.bbcm.units.it/~tossi/pag1.htm) and the AMPer site (http://marray.cmdr.ubc.ca/cgi-bin/amp.pl). Fjell *et al.*, (2007) [10] demonstrated that it is possible to accurately identify naturally-occurring host defence peptides and cluster similar peptides based solely on primary amino acid sequence using the statistical technique of hidden Markov models (HMMs). The AMPer resource has also been used to identify validated, novel host defence peptides from the genome sequence and expressed sequence (ESTs) of bovine [11]. However, these sequence analysis techniques classify naturally occurring peptides solely based on known host defence peptides in the context of all other peptides; the resulting models, though highly accurate for gene classification and identification, do not indicate the mechanisms of action of peptides or identify structural similarities between peptides. Generally, host defence peptide are typically short (12 to 50 amino acids), positively charged (net +2 to +9), amphiphilic, and can be arranged into common structural classes e.g. β-structured peptides, amphipathic α-helices, or less common loop structures and extended structures [12,13] (Fig. **1**). Despite sequence and structure variation several host defence peptides appears to demonstrate similar direct activity towards different microbes, thus making it hard to relate their structure and activity, and leaving optimization as a difficult task. Peptide design and optimization has traditionally been done based on rational sequence interpretation and screening of substitution libraries [14,15].

*Address correspondence to this author at the Roskilde University, Dept. of Science, Systems & Models, Universitetsvej 1, Building 18.1, DK-4000 Roskilde, Denmark; Tel: +45 4674 2877; Fax: +45 4674 3010; E-mail: jenssen@ruc.dk

1573-4129/10 $55.00+.00

a

Peptide	Primary amino acid sequence	PDB code	Reference
β-defensin-2	GIGDPVTC$_1$LKSGAIC$_2$HPVFC$_3$PRRYKQIGTC$_2$GLPGTKC$_1$C$_3$KKP	1FQQ	[69]
Polyphemusin-I	RRWC$_1$FRVC$_2$YRGFC$_2$YRKC$_1$R	1RKK	[70]
LL-37	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES	2K6O	[71,72]
Indolicidin	ILPWKWPWWPWRR	1G89	[73]

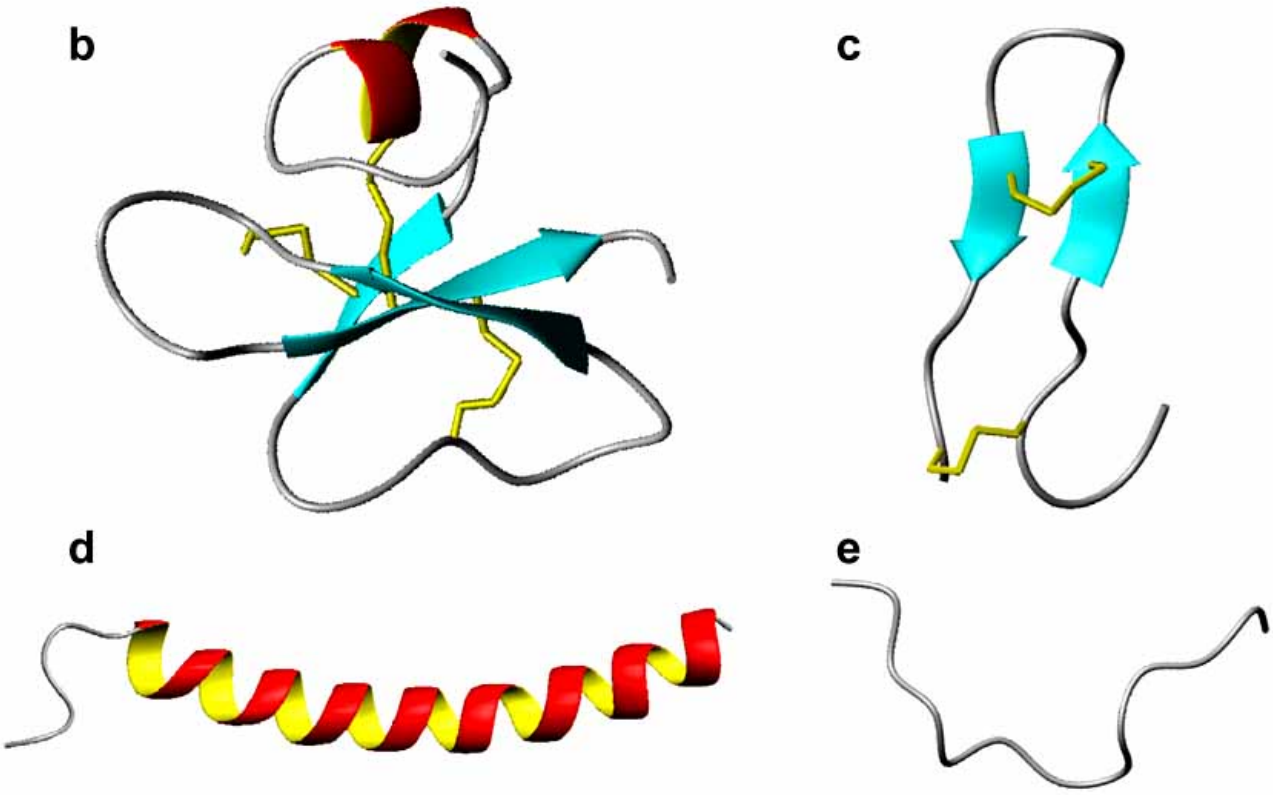

Fig. (1). Structural classes of antimicrobial peptides. (**a**) Primary amino acid sequence (one letter code) of selected host defence peptides representing the four structural classes and their PDB source code used for preparation of the crystal structure projections (**b**-**e**). Cysteines forming disulfide bonds are numbered with subscript numbers to indicate their pairing. Boldface indicates cationic (blue) and anionic (red) amino acid residues. (**b**) Mixed structure of human β-defensin-2; (**c**) β-sheeted polyphemusin; (**d**) α-helical human cathelicidin LL-37; (**e**) extended indolicidin. The peptide secondary structures are prepared with use of the graphic program MolMol 2K.2 [74] and the disulfide bonds are indicated in yellow.

Despite its success, there is no doubt that this approach is extremely labor intensive and requires insight into the peptides mode of action, which today still is highly debated. When optimizing antimicrobial peptides, new peptide candidates are synthesized based on educated guesses as to which of the molecular features are important for its activity. Chemoinformatic techniques can be used to facilitate and enforce an un-biased design of such new peptide candidates. One of the most widely used of these techniques is quantitative structure-activity relationship (QSAR) analysis. Computer-aided QSAR analysis seeks to relate quantitative properties (descriptors) of the peptide with properties such as antimicrobial activity through numerical analysis. Though the concept of QSAR is widely used in pharmaceutical drug discovery [16] the technology has more recently been implied in antimicrobial peptide design. The success of a QSAR model depends highly on the choice of QSAR descriptors and the mathematical/statistical method used to relate the descriptors to the peptides antimicrobial activity.

The descriptors can be separated in two groups: calculated descriptors such as peptide net charge and hydrophobic moment, or empirical descriptors such as HPLC retention time and measured solubility. The statistical tools used to model the peptides activity can also be divided in two groups: 'simple' regression models such as principal component analysis (PCA) and partial least squares projection to latent structures (PLS), or more advanced machine learning techniques such as artificial neural networks (ANN). This review provides an overview of some of the most common computational approaches for peptide optimization and design, and some of the most recent advancements in this field, with the emphasis on peptide antibacterial activity.

QSAR DESCRIPTORS

Descriptors are used to describe quantitative properties of the peptides and may be calculated or measured. The QSAR approach was originated by Hansch *et al.* 1962 when translating the chemical structures into numerical values describing their hydrophobicity and electronic properties [17]. Since then numerous descriptor types has been introduced to describe peptides, some more useful and intuitive than others.

There is no general rule for selecting descriptors for peptide modeling. Different types of descriptors may work equally well on different peptide sets. However, there should not be a high degree of correlation between the values of the different descriptors used. Several different calculated descriptors have been introduced. The simplest ones are probably net charge at pH 7, mean hydrophobic- or charged moment, hydrophobic fraction and Kyte-Doolittle hydrophobicity. More complex estimations are α-helical propensity calculated with equations either from Garnier, Chou-Fasman or Eisenberg [18]. The chemoinformatics advances have to some extent revolutionized descriptor generation, enabling a much more robust calculation of descriptors describing the peptides three dimensional nature. One example is the NMR grid estimation and then conversion of peptide structure into two a dimensional descriptors set [19,20]. Another example is the 'inductive' descriptors for chemical hardness, softness and steric effects, calculated based on *in silico* calculation of the most probable three dimensional structure of each peptide based on energy and flexibility constraints of the molecule [21].

Empirically derived descriptors on the other hand, are generated based on measured properties of populations of compounds in biological assays. A central aspect of peptide QSAR has been development of quantitative descriptors for the amino acids. This work was pioneered by P.H.A Sneath [22] but the real breakthrough came after Hellberg *et al.* implemented different HPLC measurements under various pH and elution conditions, and derived the z-scale [23,24]. The scale consisted of three descriptors z_1, z_2, and z_3 for each of the 20 naturally coded amino acids, describing 29 different physico-chemical properties describing lipophilicity hydrophobicity, size and charge related features of the amino acids, in addition to NMR and HPLC data. These descriptors reflect lipophilicity (z_1), steric properties (z_2), and electronic properties (z_3). The z-scale has later been refined to also describe 67 non-coded amino acids [25]. However, in modeling peptides only composed of naturally coded amino acids, the more complex and refined z-scale is not required [26].

MULTIVARIATE DATA ANALYSIS

Computer-aided models of peptide antimicrobial activity using soft independent modeling of class analogy (SIMCA), and incorporated principal component analysis (PCA)/partial least squares projection to latent structures (PLS) algorithms, have demonstrated success in explaining and modeling antimicrobial peptide activity [27-29].

In this technique an X matrix with (N) observations and (K) variables are created, using descriptor values representing the investigated peptides as well as possible. By using a multivariate projection, principal component analysis (PCA),

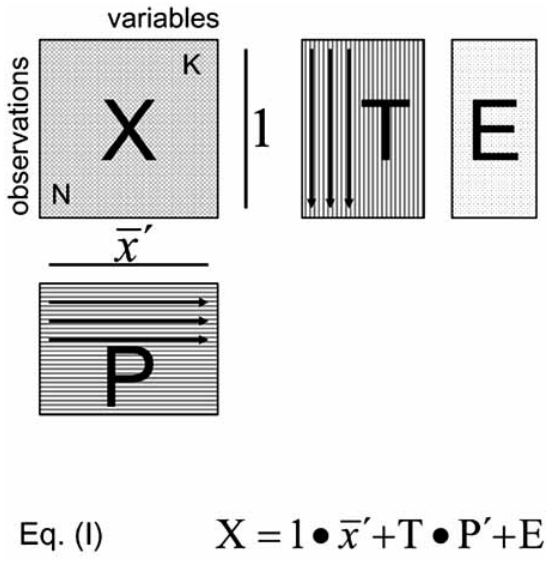

Fig. (2). Summary of PCA analysis. PCA analysis will model a table or matrix X, with N observations and K variables, as indicated in equation I. The first element ($1 \bullet \bar{x}'$) represents the average of the variables and are derived in a pre-processing step. The second element ($T \bullet P'$) models the structure while (E) represents noise in the model. The T and P matrix composed of principal component score vectors (t_1, t_2, etc.) and loading vectors (p_1, p_2, etc.), respectively.

the X matrix is converted into two matrices (T) and (P) (Fig. **2**). The score (T) describes how the tested molecules (peptides) are related to each other, while the loading (P) reveals any correlation between the variables and their importance for the PCA model [30]. PCA is performed without reference to the biological activity and linear regression is performed using a small number of dimensions (principal components) to fit the measured and predicted activity (principal component regression - PCR). PCA is a popular method to reduce the number of variables used in fitting the data, thus allowing the use of more descriptors to be included in analysis than there are samples.

Partial least squares projections to latent structures (PLS) is a method that is similar to PCR. However, in PLS the importance of each variable (i.e. each column in X) is first replaced by the column weighted by the regression coefficient of the response (Y) to that variable by itself [31]. In this way, the importance of each variable is enhanced in predicting the response. The method can handle several responses at the same time, and regression is possible even with some missing data in the matrix [32]. The validity of the model depends on how well it correlates variations in the response (Y) and in the (X) matrix (Fig. **3**). PLS strategies and more detailed descriptions of this technique have been reviewed elsewhere (see [30,33,34] for overviews).

ANTIMICROBIAL PEPTIDE

It is well understood that regardless of their actual origin and mode of action, all types of antibacterial cationic peptides must interact with the bacterial cytoplasmic membrane [35]. The physical forces behind antibacterial activity have

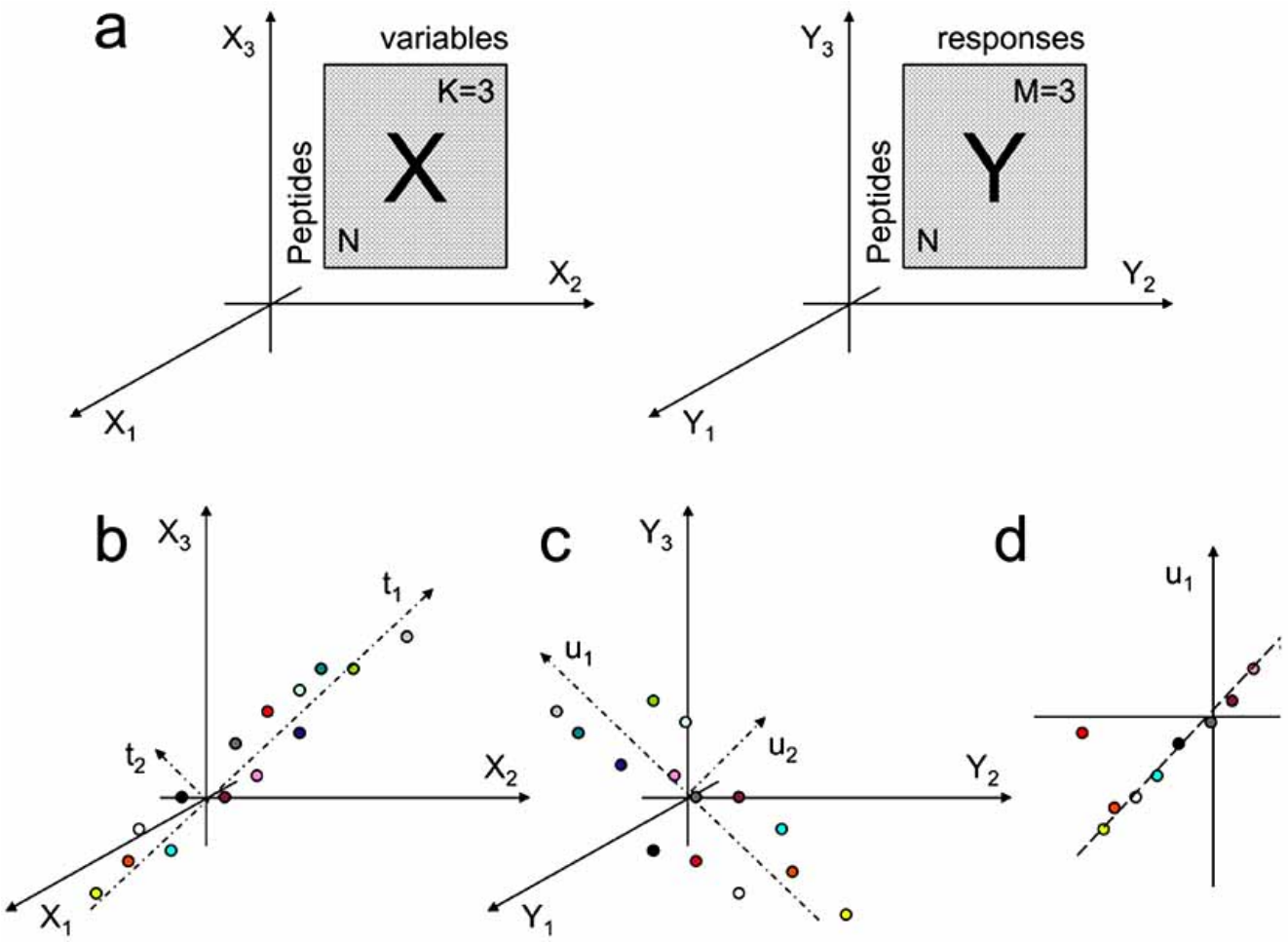

Fig. (3). Modeling of the relationship between two matrices by using PLS analysis. Matrices in PLS can be understood as fitting two "PCA-like" models, one of your X matrix and one of your Y matrix. This is illustrated for a library containing 12 peptides (color coded). The peptides have been described with three variables (descriptors) in the X matrix and three biological responses have been monitored in matrix Y. (**a**) As K and M both equal 3, the peptides in X and Y can be depicted in a three dimensional space. (**b**) The best fit line that approximates the point swarm are drawn on both graphs (t_1 and u_1), and these represents the first component in the PLS model or the "score" for X and Y, respectively. (**c**) The best fit line, drawn orthogonal on t_1 and u_1, represents the t_2 and u_2, or the second component in the PLS model. (**d**) The PLS score plot will identify the relationship between the two matrices, and plotting of t_1 against u_1 will normally correlate better than t_2 against u_2, and so on.

been defined in detail (see [35-37] for overviews) and include: net positive charge which enhance the interaction with anionic lipids and other bacterial targets, hydrophobicity allows insertion of the peptide into the bacterial membrane, and flexibility permitting the peptide to transition from its solution conformation to its membrane interacting conformation.

Each of these characteristics can vary substantially over a particular range, but are essential for the peptides function as antimicrobial agents and allows them to interact with bacterial membranes, which is critical to them exerting antimicrobial effects. The way antimicrobial peptides interact with the prokaryotic cell membrane has been extensively investigated [38-41]. Due to the fact that most antimicrobial peptides are cationic at physiological pH, they are prone to interact quite unspecifically through electrostatic interactions, with negatively charged molecules in the bacterial cell membrane. In Gram-negative bacteria these negatively charged molecules are lipopolysaccharide while in Gram-positive bacteria these molecules are lipoteichoic acid and/or teichoic acid. Experiments have demonstrated that lactoferricin can bind both lipopolysaccharide and teichoic acid [42,43].

Bovine lactoferricin (Fig. **4**) have been studies extensively over the past decades [44], and a lot of effort has been put into identify the active domain and optimizing the antimicrobial activity of this cyclic 25-residue gastric cleavage product of a milk derived protein, lactoferrin [45]. Several lactoferricin homologues from different species e.g. human, caprine, murine, procine and bovine have been investigated for their antibacterial activity, and despite rather conserved sequences, their antibacterial activity and specificity varies significantly. Bovine lactoferricin has superior bactericidal activity compared to the other lactoferricin peptides [46,47]. Gram-positive bacteria have also been shown more susceptible to lactoferricin than Gram-negative bacteria, probably due to the lack of an outer membrane [48].

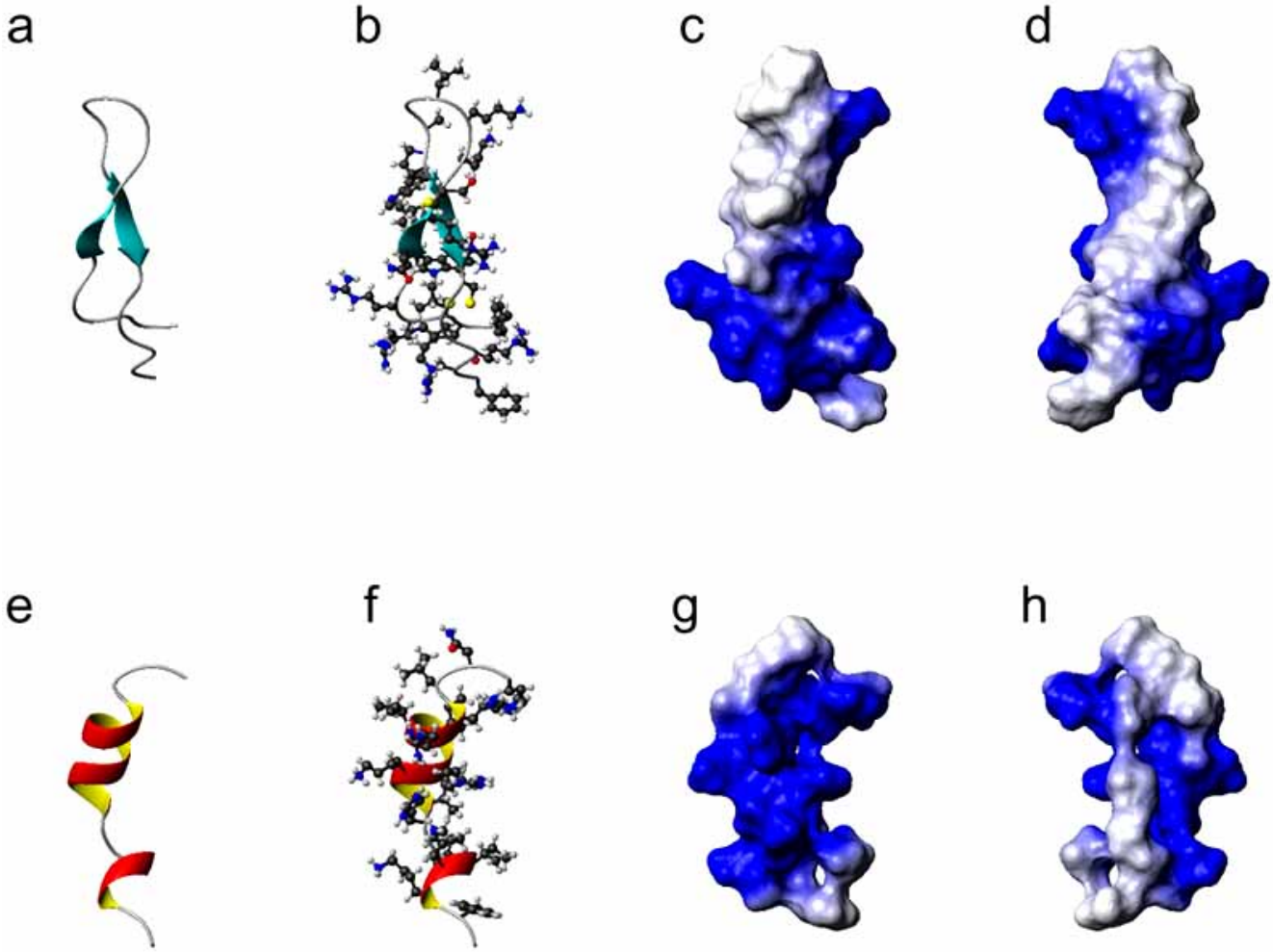

Fig. (4). Peptide crystal structures peptides. (**a-d**) β-sheet bovine lactoferricin (PDB code 1LFC) [75] and (**e-h**) α-helical novispirin (PDB code 1HU6) [63] are shown. (**a** and **b**) illustrates the plain ribbon structure (blue β-sheet) of lactoferricin in the absence and presence of the reactive amino acid side chains, respectively. (**c**) A charge distribution plot of lactoferricin in the same orientation as (**a**), colored blue and white corresponding to net positive and neutral charge, respectively. (**d**) Charge distribution plot of lactoferricin from diagram (**c**) rotated 180° around the Y-axis. (**e** and **f**) illustrates the plain ribbon structure (red α-helix) of novispirin in the absence and presence of the reactive amino acid side chains, respectively. (**g**) A charge distribution plot of novispirin in the same orientation as (**f**), colored blue and white corresponding to net positive and neutral charge, respectively. (**h**) Charge distribution plot of novispirin from diagram (**g**) rotated 180° around the Y-axis. The figure was prepared with use of the graphic program MolMol 2K.2 [74].

However the results also demonstrate that despite very different secondary peptide structure, both bovine and human lactoferricin are able to exert antibacterial activity, indicating that antibacterial activity and spectrum of a peptide not can be extrapolated from the peptides secondary structure [7,49], but rather their amphipathic and amphiphilic patches in their folded structure and by regions with high concentration of positively charged residues [50]. However, modulation of the antibacterial activity of cationic peptide through alteration of their hyrdophobicity or net charge may also alter the selectivity between the desired bacterial target and the host cell [51,52]. Similarly, incorporation of charged residues above a certain maximum (varying from peptide to peptide) does not lead to an increase in activity [36]. Thus this balance of charge and hydrophobicity can be delicate and must be empirically determined for each series of peptides.

In an attempt to better understand the underlying mechanisms of the antibacterial activity of lactoferricin, principal component analysis (PCA) was introduced. As mentioned earlier, the descriptors used in this multivariate data analysis approach determine the information gained from the modeling. Consequently a comparative study was conducted investigating the structural requirements for antibacterial activity of murine lactoferricin, using two separate sets of descriptors, i.e. one model using empirically derived amino acid descriptors [24] and one model using calculated descriptors e.g. charge, helicity factors and propensity [18,28]. The study revealed no significant difference amongst the two datasets, and the respective models. For the sake of modeling peptide abilities as antimicrobials, it was judged that describing the peptide building blocks (amino acids) with the z-scale descriptors introduced by Hellberg *et al.* [24], rather that calculating the physiochemical property of the entire

peptide molecule would be better for the purpose of modeling. QSAR analysis of peptides using the z-scale descriptors in PCA modeling has later been proven effective in explaining and predicting a spectrum of different antimicrobial activities e.g. antibacterial [53], anti-cancer [27], as well as antiviral activity [29].

In all these experiments a limited set of lactoferricin derived peptides were investigated for a limited set of biological activities. A pressing concern where the structural restrains amongst the peptides modeled. To verify that the PCA model also would work on peptides with different structural motifs, a set of α-helical peptides where examined. This study demonstrated that both peptide antiviral activity as well as peptide interactions with cell surface receptor molecules could be understood with the computational algorithms [54]. The general conclusion was that peptide charge- and hydrophobicity-related properties demonstrated the highest importance in describing variation among the peptides. These results introduced a glimmer of hope for peptide activity optimization *in silico*, and to test the potential of this partial least squares projections to latent structures (PLS) model a vast set of virtual peptide sequences were designed. The antiviral activity of a total of ~218.000 virtual peptide sequences were predicted with this model. The success of the prediction was evaluated by synthesizing a limited set of the peptides predicted to have the most potent antiviral activity. The results revealed half the peptides predicted to be highly active, were 2.7 to 4.6 fold more active than the most active peptide in the model; on the other hand, half the peptides predicted to be highly active were found to be inactive [55]. This demonstrates a fundamental problem with the PLS model. The model is based on a linear relationship, and though the model may have a close to perfect relation ship between the peptides incorporated, placement of virtual peptide may be problematic if they differ too much from the model peptides. It can easily be interpreted that the peptide is significantly better or worse than the model peptides, however how much better or worse is harder to predict. Consequently, the accuracy of the predictive model will drop as the peptides are significantly better or worse than the ones in the model. Though it is claimed that construction of antimicrobial peptides using multivariate design will solve the problem of introducing multiple amino acid changes during peptide synthesis [56], there is no doubt that multiple substitutions in a single peptide also to some extent will make it harder to precisely predict its activity [53].

On the other hand, single substitutions may also cause critical problems for PLS modeling of antimicrobial peptides. Features recognized as essential with direct antimicrobial peptides have for a long time been their high content of charged and hydrophobic amino acids. When optimizing these peptides through single substitutions there is a likelihood of generating several peptides with different primary sequences, but the same amino acid content. However, a persistent problem with modeling efforts has been that no primary structure information has been implemented in the models. Thus, modeling of such peptides, using only the specific amino acid descriptors (z-scale) will result in a model interpreting the peptides as identical, consequently making it close to impossible to make good predictions. In an attempt to circumvent this problem contact energy between neighboring amino acids [57] was introduce as a descriptor, for modeling >200 peptides from a single substitution library of the 12mer peptide, Bac2A (RLARIVVIR-VAR-NH_2) [15]. Although this ignores all intramolecular interactions involved in determining three dimensional structure, except for those between neighboring amino acids in the primary structure, this implementation resulted in a rather powerful predictive model [26].

It should be pointed out that the contact energy values between neighboring amino acids have been derived for a small set of proteins [57]. Thus using them as descriptors for design of PLS models for antimicrobial peptides may be oversimplified, and in some cases it may yield poor predictive performance. Another aspect is that in this particular study it was demonstrated that the contact energy descriptors in combination with a subset of 'inductive' and conventional QSAR descriptors [21] gave a more robust model than by using the contact energy descriptors by itself [26]. It was also confirmed in a consecutive study that use of the contact energy descriptors also gave a relatively good predictive model for a different set of peptides [58].

Although this is the first time two distinct subgroups of peptides successfully has been used to generate a predictive model, though with inconsistent accuracy, it is well known that models based on different descriptors will yield different results. A classical example is from one of the earliest comparative PCA studies on mouse lactoferricin that was mentioned earlier. In this study two different sets of descriptors were used and proven to give very similar models [18,28]. In this work a set of murine lactoferricin derivatives where tested together with bovine lactoferricin. The murine derivatives and the bovine peptide are significantly different from a primary sequence point of view, when modeled with standard amino acid descriptors (z-scale). This is also indirectly confirmed by Lejon *et al.* when they with high accuracy are able to correlate predicted and observed antibacterial activity of the murine derivatives, while the bovine peptide is very poorly modeled [28]. On the other hand when examining secondary structure resemblance between lactoferricin peptides from other origin [59], it can be assumed that there is a rather striking resemblance between the bovine and the murine peptides in terms their secondary structures. When using calculated descriptors more related to the secondary structural properties of the entire peptide instead of focusing on independent amino acids, the PCA model demonstrated a greater success in correlating predicted and observed antibacterial peptide activity of all the peptides, including bovine lactoferricin [18].

The 'inductive' and conventional QSAR descriptors are computer simulated parameters describing biophysical properties of the entire peptide [21] and some more commonly-used QSAR-descriptors as implemented in MOE (*Molecular Operating Environment* software v. 2006.10, Chemical Computation Group Inc., Montreal, Canada, 2006). These inductive and conventional QSAR descriptors has prior to the demonstrated usefulness in PLS modeling of antibacterial peptides [26,58] also demonstrated tremendous success when used in combination with artificial intelligence approach for predicting antimicrobial activity of a limited set of organic molecules [60,61]. The great success of using these descrip-

tors in PLS modeling of antimicrobial peptides is likely due to the descriptors' ability to capture three dimensional structural differences amongst the peptides. Use of the z-scale descriptors alone will focus the PLS model around specific changes at an amino acid level, and most probably miss out on steric and intermolecular changes in the peptide as a hole. While the computers simulation of the peptides three dimensional structure, based on energy levels and folding potential will give a better understanding of all parts of the peptides. However, it should be kept in mind that peptides are intrinsically flexible, and may adopt several 'stable' conformations in different environments and upon interaction with different cellular components and membranes. Thus, one may argue for a potential in improving even the 'inductive' and conventional QSAR descriptor set.

Another very robust PLS model has been demonstrated for a totally different class of peptides. The α-helical sheep myeloid antimicrobial peptide (SMAP-29) is a cathelicidin-derived antimicrobial peptide from leukocytes [62] (Fig. **4**). Preliminary sequence optimization and QSAR work on this peptide led to the generation of ovispirin-1 which has high resemblance to the N-terminal part of SAMP-29, and to the less cytotoxic analogue novispirin G_{10} (KNLRRI-IRKG$_{10}$IHIIKKYG) with a single glycine substitution in position ten [63], Novispirin G_{10} has later served as a template in a thorough QSAR study using PCA/PLS modeling [64]. In this study a total of 58 novispirin analogues where synthesized, and a predictive model were generated using 69 molecular descriptors i.e. Vol-Surf (http://www.moldiscovery.com) and charged partial surface area descriptors, being most important. The model was used to predict antibacterial activity of 400 virtual peptides, and the success of the model was evaluated by synthesizing and testing sixteen of the peptides predicted as highly active. The model demonstrated a 75% success rate in predicting highly active peptides, with three out of four peptides tested being more active than the parent novispirin G_{10}. This model was generated without using standard amino acid descriptors (z-scale), and the robustness of it is probably due to introduction of descriptors dealing with three dimensional structural parameters of the entire peptide.

MACHINE LEARNING AND NEURAL NETWORK MODELING

A large literature has been developed on the complex modeling techniques grouped together as machine learning. Detailed description of this large field of study is outside the scope of this review; however we will highlight recent work that successfully applied artificial neural networks (ANNs) to the prediction of highly-active synthetic peptides. ANNs are one of the oldest machine learning methods and relies on a 'black-box' approach: peptide data in the form of both descriptors and measured activity are provided and a complex algorithm attempts to extract a pattern from these 'training data'. In an ANN, the descriptor values are applied to the input layer of the network. The calculation propagates from the input layer to the first of possibly many hidden layers that are 'connected' to the input layer: the inputs to the hidden layers are weights sums of the input values. The output of each hidden node is a non-linear transformation of its input which is then passed in a weighted sum to the next layer. Finally the values are passed to the output layers consisting of one or more nodes.

ANN was introduced in antimicrobial peptide modeling more than a decade ago. This preliminary study was rather limited and did not result in overwhelming results for either antimicrobial activity nor predictive power [65], and most-likely escaped the attention of most microbiologists. In recent peptide prediction work using semi-random 9-mer peptides [66,67], the ANN configuration used 44 input nodes (one per descriptor), one hidden layer of 10 nodes, and one output node that represented the prediction of the peptide being active or inactive (Fig. **5a**). The output from a node are afunction of the weighted sum of the inputs plus a bias. During training, the weighting values, represented by lines connecting nodes, are systematically altered to minimize the disagreement between the given measured activity (1 for active, 0 for inactive) and the predicted activity (a value between 0 and 1) (Fig. **5b**). Thus, for this study, over 450 weighting parameters (there are several offset parameters not represented in the figure) were required to be fit for each network. Since some fraction of peptides must be held out of the training set for validation of the model, data from more than 450 peptides are required for this method to work. Once the networks are trained, sets of descriptor values are given as input and the output node indicates the prediction of activity.

The requirement for a large number of samples for training is a general feature of complex models and cannot be avoided without loss of predictive power. This represents the greatest drawback of using complex models: for many studies the number of data required is simply not available.

COMPUTATIONAL METHODS – ADVANTAGES AND LIMITATIONS

The most commonly used computational QSAR approach is probably principal component analysis (PCA) and the closely related method, projections to latent structures (PLS). This is probably due to the nature of this mathematical approach. The technique handles missing values (incomplete libraries/data sets) quite well. It is also possible to use this approach to generate meaningful models on rather restricted (small) data sets. The technique does neither require 'super computers' which made this technology possible for a wide audience, even back in the days when computers where looked upon as an expensive tool rather than a consumable. The technique is also excellent in identifying outliers, or sub-groups with similar activity patterns. The relatively 'simple' nature of this statistical approach makes it possible to evaluate the models predictive ability and answering why some candidates are predicted to be better than others. However, there are limitations to the predictive ability of these models. Any peptide candidate predicted to be significantly better than the most active peptide in the model, will in theory be assigned a value far below the most active peptide (the lower the value the higher activity). This is however only theoretical, and peptides which are predicted to lie far outside the model will have to some extent an equal chance of being predicted as super-active or inactive. This phenomenon has been demonstrated for a set of α-helical peptides designed to inhibit herpes simplex virus infection of

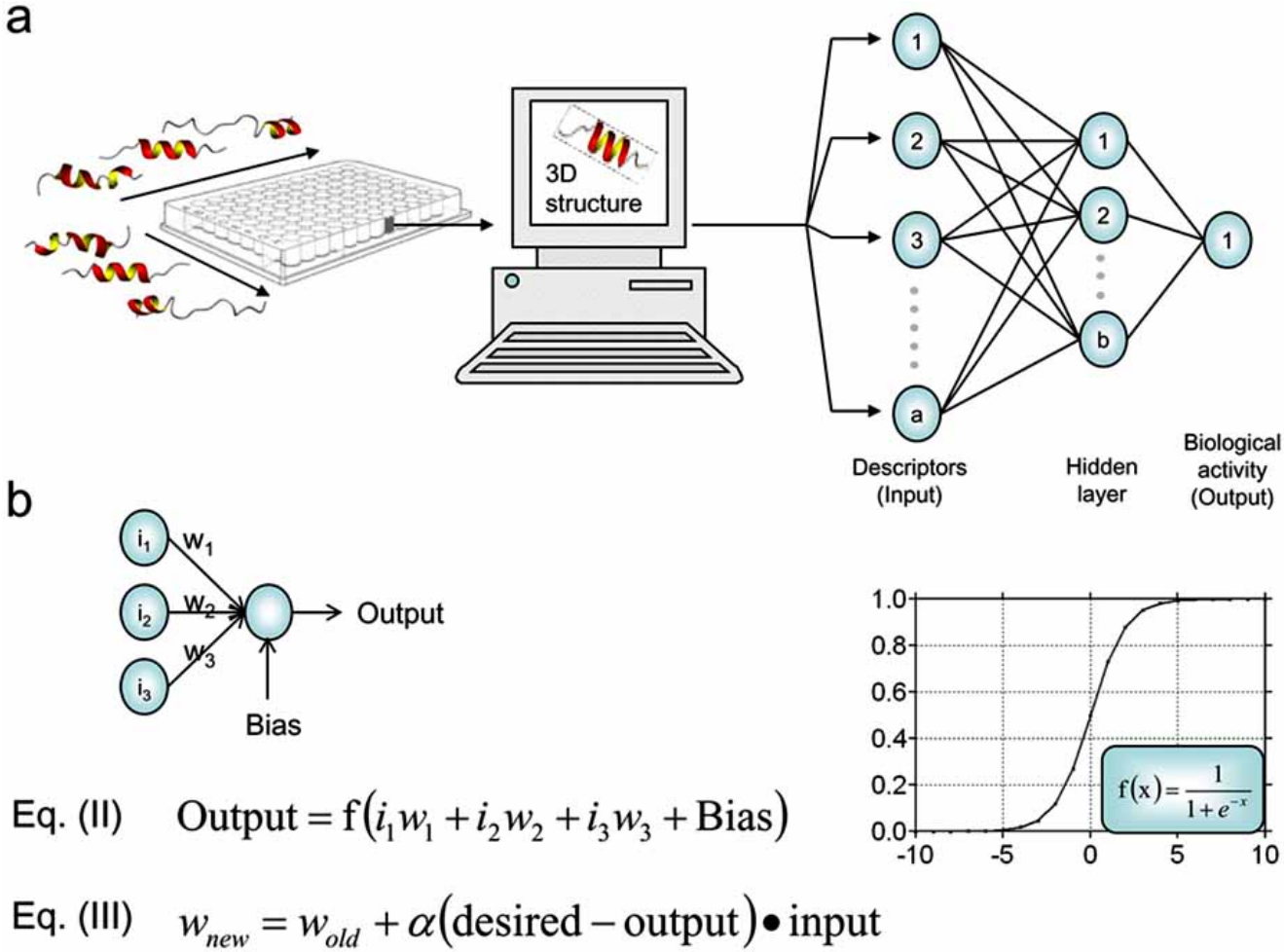

Fig. (5). Flow chart of peptide optimization using artificial neural network networks. The use of an artificial neural network in peptide antimicrobial activity optimization was described by Cherkasov *et al.* [66,67]. (**a**) The three dimentional structure of all the peptides in the library were simulated in gas phase using energy minimization restrictions and the generalized Born solvation model. Chemical descriptors where calculated using MOE (Molecular Operating Environment, 2005, Chemical Computing Group Inc., Montreal, Canada), taking into account all of the atoms in the peptides. The calculated descriptors are then feed into the neural network. The specific network in the reviewed example consisted of three layers: the input layer with 44 normalized QSAR descriptor values (a=44), a hidden layer (may have more than one) with 10 nodes (b=10) and the output layer which contains the biological readouts (antibacterial activity) was monitored (c=1). (**b**) Calculating the weights of the different interactions was done using equation (II), and the sigmoidal learning function demonstrates the non linear transformation between input and outputs. The learning process or training of the network was done using equation (III), where the weights are adjusted so that the neural network output value better matches the biologically measured activity.

host cells. Where a PLS model was used to predict the activity of ~218.000 peptides, and when evaluating the results with synthesis of the most active candidates, half of them were evaluated as more active than the most active peptide in the model, and the other half was measured as inactive [55]. Though descriptor generation continuously is evolving there is currently no optimal descriptor set that will enable highly accurate modeling of peptides with large structural diversity using PLS modeling [58]. Thus the *in silico* screening of large libraries and optimization or design of truly novel peptide candidates based on a single library of peptides is difficult using PLS.

Non-linear techniques and artificial neural networks (ANNs) (Fig. **5**) are much more complex than PLS models and are therefore more able to capture patterns indicating peptide activity. Consequently they may give superior results, with the important requirement that data from large numbers of peptides are available. However this improved performance is also gained at the cost of the rather cryptic nature of the models: their complexity makes it difficult to relate the contribution of each descriptor to the activity of the peptides [68].

ACKNOWLEDGEMENT

We gratefully acknowledge financial support through grants from the Canadian Institutes for Health Research (CIHR) and the Applied Food and Materials Network. REWH was the recipient of a Canada Research Chair.

REFERENCES

[1] Moran, G.J.; Mount, J., Update on emerging infections: news from the Centers for Disease Control and Prevention. *Ann. Emerg. Med.*, **2003**, *41*(1), 148-151.

[2] Wenzel, R.P. The antibiotic pipeline-challenges, costs, and values. *N. Engl. J. Med.*, **2004**, *351*(6), 523-526.

[3] Norrby, S.R. Alert to a European epidemic. *Nature*, **2004**, *431*(7008), 507-508.

[4] Projan, S.J. Why is big Pharma getting out of antibacterial drug discovery? *Curr. Opin. Microbiol.*, **2003**, *6*(5), 427-430.

[5] Chan, P.F.; Holmes, D.J.; Payne, D.J. Finding the gems using genomic discovery: antibacterial drug discovery strategies – the successes and the challenges. DDT: *Ther. Strategies.*, **2004**, *1*, 519-527.

[6] Hancock, R.E.W.; Sahl, H.G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.*, **2006**, *24*(12), 1551-1557.

[7] Jenssen, H.; Hamill, P.; Hancock, R.E.W. Peptide antimicrobial agents. *Clin. Microbiol. Rev.*, **2006**, *19*(3), 491-511.

[8] Finlay, B.B.; Hancock, R.E.W. Can innate immunity be enhanced to treat microbial infections? *Nat. Rev. Microbiol.*, **2004**, *2*(6), 497-504.

[9] Hancock, R.E.W. Cationic peptides: effectors in innate immunity and novel antimicrobials. *Lancet Infect. Dis.*, **2001**, *1*(3), 156-164.

[10] Fjell, C.D.; Hancock, R.E.W.; Cherkasov, A. AMPer: a database and an automated discovery tool for antimicrobial peptides. *Bioinformatics*, **2007**, *23*(9), 1148-1155.

[11] Fjell, C.D.; Jenssen, H.; Fries, P.; Aich, P.; Griebel, P.; Hilpert, K.; Hancock, R.E.W.; Cherkasov, A. Identification of novel host defense peptides and the absence of alpha-defensins in the bovine genome. *Proteins*, **2008**, *73*, 420-430.

[12] Bowdish, D.M.; Davidson, D.J.; Hancock, R.E.W. A re-evaluation of the role of host defence peptides in mammalian immunity. *Curr. Protein Pept. Sci.*, **2005**, *6*(1), 35-51.

[13] Boman, H.G. Peptide antibiotics and their role in innate immunity. *Annu. Rev. Immunol.*, **1995**, *13*, 61-92.

[14] Hilpert, K.; Elliott, M.R.; Volkmer-Engert, R.; Henklein, P.; Donini, O.; Zhou, Q.; Winkler, D.F.; Hancock, R.E.W. Sequence requirements and an optimization strategy for short antimicrobial peptides. *Chem. Biol.*, **2006**, *13*(10), 1101-1107.

[15] Hilpert, K.; Volkmer-Engert, R.; Walter, T.; Hancock, R.E.W. High-throughput generation of small antibacterial peptides with improved activity. *Nat. Biotechnol.*, **2005**, *23*(8), 1008-1012.

[16] Perkins, R.; Fang, H.; Tong, W.; Welsh, W.J. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.*, **2003**, *22*(8), 1666-1679.

[17] Hancsh, C.; Maloney, P.P.; Fujita, T.; Muri, R.M. Correlation of biological activity of phenozyacetic acid with hammett substituent constants and partition coefficients. *Nature*, **1962**, *194*, 178-180.

[18] Strom, M.B.; Rekdal, O.; Stensen, W.; Svendsen, J.S. Increased antibacterial activity of 15-residue murine lactoferricin derivatives. *J. Pept. Res.*, **2001**, *57*(2), 127-139.

[19] Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.*, **2000**, *11*(Suppl 2), S29-39.

[20] Zamora, I.; Oprea, T.; Cruciani, G.; Pastor, M.; Ungell, A.L. Surface descriptors for protein-ligand affinity prediction. *J. Med. Chem.*, **2003**, *46*(1), 25-33.

[21] Cherkasov, A. Inductive QSAR descriptors, distinguishing compounds with antibacterial activity by artificial neural network. *Int. J. Mol. Sci.*, **2005**, *6*, 63-86.

[22] Sneath, P.H. Relations between chemical structure and biological activity in peptides. *J. Theor. Biol.*, **1966**, *12*(2), 157-195.

[23] Hellberg, S.; Sjostrom, M.; Wold, S. The prediction of bradykinin potentiating potency of pentapeptides. An example of a peptide quantitative structure-activity relationship. *Acta. Chem. Scand. B*, **1986**, *40*(2), 135-140.

[24] Hellberg, S.; Sjostrom, M.; Skagerberg, B.; Wold, S., Peptide quantitative structure-activity relationships, a multivariate approach. *J. Med. Chem.*, **1987**, *30*(7), 1126-35.

[25] Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjostrom, M.; Wold, S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J. Med. Chem.*, **1998**, *41*(14), 2481-2491.

[26] Jenssen, H.; Lejon, T.; Hilpert, K.; Fjell, C.D.; Cherkasov, A.; Hancock, R.E.W. Evaluating different descriptors for model design of antimicrobial peptides with enhanced activity toward P. aeruginosa. *Chem. Biol. Drug. Des.*, **2007**, *70*(2), 134-142.

[27] Yang, N.; Lejon, T.; Rekdal, O. Antitumour activity and specificity as a function of substitutions in the lipophilic sector of helical lactoferrin-derived peptide. *J. Pept. Sci.*, **2003**, *9*(5), 300-311.

[28] Lejon, T.; Strom, M.B.; Svendsen, J.S. Antibiotic activity of pentadecapeptides modelled from amino acid descriptors. *J. Pept. Sci.*, **2001**, *7*(2), 74-81.

[29] Jenssen, H.; Gutteberg, T.J.; Lejon, T. Modelling of anti-HSV activity of lactoferricin analogues using amino acid descriptors. *J. Pept. Sci.*, **2005**, *11*(2), 97-103.

[30] Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Trygg, J.; Wikström, C.; Wold, S. *Multi- and Megavariate Data Analysis - Basic Principles and Applications*. Umetrics: Umeå **2006**, Vol. *1*, pp. 1-425.

[31] Hastie, T.; Tibshirani, R.; Friedman, J. *The elements of statistical learning data mining, inference, and prediction.* Springer: New York **2001**, pp. 66-67.

[32] Eriksson, L.; Jonsson, J.; Hellberg, S.; Lindgren, F.; Skagerberg, B.; Sjostrom, M.; Wold, S. Peptide QSAR on substance P analogues, enkephalins and bradykinins containing L- and D-amino acids. *Acta. Chem. Scand.*, **1990**, *44*(1), 50-55.

[33] Wold, S.; Albano, C.; Dunn, W.J., III.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. In: *Chemometrics - mathematics and statistics in chemistry*; Kowalski, B.R., Ed.; D. Reidel Publishing Company: Dordrecht **1984**, Vol. 138, pp. 17-95.

[34] Wold, S.; Sjöström, M., In: *Chemometrics: theory and application*, Kowalski, B. R., Ed.; American Chemical Society: washington, **1977**; Vol. 52, pp. 243-282.

[35] Hancock, R.E.W.; Rozek, A. Role of membranes in the activities of antimicrobial cationic peptides. *FEMS. Microbiol. Lett.*, **2002**, *206*(2), 143-149.

[36] Dathe, M.; Wieprecht, T. Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. *Biochim. Biophys. Acta.*, **1999**, *1462*(1-2), 71-87.

[37] Hancock, R.E.W.; Patrzykat, A. Clinical development of cationic antimicrobial peptides: from natural to novel antibiotics. *Curr. Drug Targets Infect. Disord.*, **2002**, *2*(1), 79-83.

[38] Andreu, D.; Rivas, L. Animal antimicrobial peptides: an overview. *Biopolymers*, **1998**, *47*(6), 415-433.

[39] Epand, R.M.; Vogel, H.J. Diversity of antimicrobial peptides and their mechanisms of action. *Biochim. Biophys. Acta.*, **1999**, *1462*(1-2), 11-28.

[40] Matsuzaki, K.; Sugishita, K.; Ishibe, N.; Ueha, M.; Nakata, S.; Miyajima, K.; Epand, R.M. Relationship of membrane curvature to the formation of pores by magainin 2. *Biochemistry*, **1998**, *37*(34), 11856-11863.

[41] Sitaram, N.; Nagaraj, R. Interaction of antimicrobial peptides with biological and model membranes: structural and charge requirements for activity. *Biochim. Biophys. Acta.*, **1999**, *1462*(1-2), 29-54.

[42] Chapple, D.S.; Mason, D.J.; Joannou, C.L.; Odell, E.W.; Gant, V.; Evans, R.W. Structure-function relationship of antibacterial synthetic peptides homologous to a helical surface region on human lactoferrin against *Escherichia coli* serotype O111. *Infect. Immun.* **1998**, *66*(6), 2434-2440.

[43] Vorland, L.H.; Ulvatne, H.; Rekdal, O.; Svendsen, J.S. Initial binding sites of antimicrobial peptides in Staphylococcus aureus and *Escherichia coli*. *Scand. J. Infect. Dis.*, **1999**, *31*(5), 467-473.

[44] Faber, C.; Stallmann, H.P.; Lyaruu, D.M.; Joosten, U.; von Eiff, C.; van Nieuw Amerongen, A.; Wuisman, P.I. Comparable efficacies of the antimicrobial peptide human lactoferrin 1-11 and gentamicin in a chronic methicillin-resistant *Staphylococcus aureus* osteomyelitis model. *Antimicrob. Agents Chemother.*, **2005**, *49*(6), 2438-2444.

[45] Bellamy, W.; Takase, M.; Yamauchi, K.; Wakabayashi, H.; Kawase, K.; Tomita, M. Identification of the bactericidal domain of lactoferrin. *Biochim. Biophys. Acta .*, **1992**, *1121*(1-2), 130-136.

[46] Strom, M.B.; Haug, B.E.; Rekdal, O.; Skar, M.L.; Stensen, W.; Svendsen, J.S. Important structural features of 15-residue lactoferricin derivatives and methods for improvement of antimicrobial activity. *Biochem. Cell Biol.*, **2002**, *80*(1), 65-74.

[47] Vorland, L.H.; Ulvatne, H.; Andersen, J.; Haukland, H.; Rekdal, O.; Svendsen, J. S.; Gutteberg, T.J., Lactoferricin of bovine origin is more active than lactoferricins of human, murine and caprine origin. *Scand. J. Infect. Dis.*, **1998**, *30*(5), 513-517.

[48] Ulvatne, H.; Samuelsen, O.; Haukland, H.H.; Kramer, M.; Vorland, L.H. Lactoferricin B inhibits bacterial macromolecular synthesis in

Escherichia coli and Bacillus subtilis. *FEMS. Microbiol. Lett.*, **2004**, *237*(2), 377-384.

[49] Friedrich, C.L.; Moyles, D.; Beveridge, T.J.; Hancock, R.E.W., Antibacterial action of structurally diverse cationic peptides on gram-positive bacteria. *Antimicrob. Agents Chemother.*, **2000**, *44*(8), 2086-2092.

[50] Powers, J.P.; Rozek, A.; Hancock, R.E.W. Structure-activity relationships for the beta-hairpin cationic antimicrobial peptide polyphemusin I. *Biochim. Biophys. Acta.*, **2004**, *1698*(2), 239-250.

[51] Kustanovich, I.; Shalev, D.E.; Mikhlin, M.; Gaidukov, L.; Mor, A. Structural requirements for potent versus selective cytotoxicity for antimicrobial dermaseptin S4 derivatives. *J. Biol. Chem.*, **2002**, *277*(19), 16941-16951.

[52] Zelezetsky, I.; Pag, U.; Sahl, H.G.; Tossi, A. Tuning the biological properties of amphipathic alpha-helical antimicrobial peptides: rational use of minimal amino acid substitutions. *Peptides*, **2005**, *26*(12), 2368-2376.

[53] Lejon, T.; Stiberg, T.; Strom, M.B.; Svendsen, J.S. Prediction of antibiotic activity and synthesis of new pentadecapeptides based on lactoferricins. *J. Pept. Sci.*, **2004**, *10*(6), 329-335.

[54] Jenssen, H.; Gutteberg, T.J.; Lejon, T. Modelling the anti-herpes simplex virus activity of small cationic peptides using amino acid descriptors. *J. Pept. Res.*, **2005**, *66* (Suppl 1), 48-56.

[55] Jenssen, H.; Gutteberg, T.J.; Rekdal, O.; Lejon, T. Prediction of activity, synthesis and biological testing of anti-HSV active peptides. *Chem. Biol. Drug Des.*, **2006**, *68*(1), 58-66.

[56] Sjöström, M.; Eriksson, L. In *Chemometircs methods in molecular design*, Waterbeemd, H.V.D., Ed.; VCH Verlagsgesellshaft mbH: Weinheim **1995**, pp. 63-90.

[57] Miyazawa, S.; Jernigan, R.L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, **1996**, *256*(3), 623-644.

[58] Jenssen, H.; Fjell, C.D.; Cherkasov, A.; Hancock, R.E.W. QSAR modeling and computer-aided design of antimicrobial peptides. *J. Pept. Sci.*, **2008**, *14*(1), 110-114.

[59] Jenssen, H. Anti herpes simplex virus activity of lactoferrin/lactoferricin - an example of antiviral activity of antimicrobial protein/peptide. *Cell Mol. Life Sci.*, **2005**, *62*(24), 3002-3013.

[60] Cherkasov, A. Can 'Bacterial-Metabolite-Likeness' model improve odds of 'in silico' antibiotic discovery? *J. Chem. Inf. Model*, **2006**, *46*(3), 1214-1222.

[61] Karakoc, E.; Sahinalp, S.C.; Cherkasov, A. Comparative QSAR-and fragments distribution analysis of drugs, druglikes, metabolic substances, and antimicrobial compounds. *J. Chem. Inf. Model*, **2006**, *46*(5), 2167-2182.

[62] Skerlavaj, B.; Benincasa, M.; Risso, A.; Zanetti, M.; Gennaro, R. SMAP-29: a potent antibacterial and antifungal peptide from sheep leukocytes. *FEBS. Lett.*, **1999**, *463*(1-2), 58-62.

[63] Sawai, M.V.; Waring, A.J.; Kearney, W.R.; McCray, P.B., Jr.; Forsyth, W.R.; Lehrer, R.I.; Tack, B.F. Impact of single-residue mutations on the structure and function of ovispirin/novispirin antimicrobial peptides. *Protein Eng.*, **2002**, *15*(3), 225-232.

[64] Taboureau, O.; Olsen, O.H.; Nielsen, J.D.; Raventos, D.; Mygind, P.H.; Kristensen, H.H. Design of novispirin antimicrobial peptides by quantitative structure-activity relationship. *Chem. Biol. Drug Des.*, **2006**, *68*(1), 48-57.

[65] Patel, S.; Stott, I.P.; Bhakoo, M.; Elliott, P. Patenting computer-designed peptides. *J. Comput. Aided. Mol. Des.*, **1998**, *12*(6), 543-56.

[66] Cherkasov, A.; Hilpert, K.; Jenssen, H.; Fjell, C.D.; Waldbrook, M.; Mullaly, S.C.; Volkmer, R.; Hancock, R.E.W. Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. *ACS Chem. Biol.*, **2009**, *4*(1), 65-74.

[67] Fjell, C.D.; Jenssen, H.; Hilpert, K.; Cheung, W.A.; Panté, N.; Hancock, R.E.W.; Cherkasov, A. Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem.*, **2009**, *52*(2), 2006-2015.

[68] Weaver, D.C. Applying data mining techniques to library design, lead generation and lead optimization. *Curr. Opin. Chem. Biol.*, **2004**, *8*(3), 264-270.

[69] Sawai, M.V.; Jia, H.P.; Liu, L.; Aseyev, V.; Wiencek, J.M.; McCray, P.B., Jr.; Ganz, T.; Kearney, W.R.; Tack, B.F. The NMR structure of human beta-defensin-2 reveals a novel alpha-helical segment. *Biochemistry*, **2001**, *40*(13), 3810-3816.

[70] Powers, J.P.; Tan, A.; Ramamoorthy, A.; Hancock, R.E.W. Solution structure and interaction of the antimicrobial polyphemusins with lipid membranes. *Biochemistry*, **2005**, *44*(47), 15504-15513.

[71] Gudmundsson, G.H.; Agerberth, B.; Odeberg, J.; Bergman, T.; Olsson, B.; Salcedo, R. The human gene FALL39 and processing of the cathelin precursor to the antibacterial peptide LL-37 in granulocytes. *Eur. J. Biochem.*, **1996**, *238*(2), 325-332.

[72] Wang, G. Structures of human host defense cathelicidin LL-37 and its smallest antimicrobial peptide KR-12 in lipid micelles. *J. Biol. Chem.*, **2008**, *283*(47), 32637-32643.

[73] Rozek, A.; Friedrich, C.L.; Hancock, R.E.W., Structure of the bovine antimicrobial peptide indolicidin bound to dodecylphosphocholine and sodium dodecyl sulfate micelles. *Biochemistry*, **2000**, *39*(51), 15765-15774.

[74] Koradi, R.; Billeter, M.; Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **1996**, *14*(1), 29-32.

[75] Hwang, P.M.; Zhou, N.; Shan, X.; Arrowsmith, C.H.; Vogel, H.J., Three-dimensional solution structure of lactoferricin B, an antimicrobial peptide derived from bovine lactoferrin. *Biochemistry*, **1998**, *37*(12), 4288-4298.

Received: 05 November, 2009 Revised: 18 December, 2009 Accepted: 11 February, 2010